A Pseudo-Isometric Embedding Algorithm*

David W. Dreisigmeyer†
Los Alamos National Laboratory
Los Alamos, NM 87544

and

Michael Kirby‡
Department of Mathematics
Colorado State University
Fort Collins, CO 80523

February 27, 2007

Abstract

We present a new Whitney-like algorithm for finding a low-dimensional pseudo-isometric embedding of a sampled Riemannian manifold. Tangent spaces on the manifold are estimated from the data and then projected using a criterion that ensures optimal smoothness of the inverse. This short projection is not isometric but can be made to be approximately isometric by determining an appropriate global lengthening transformation in the embedded space. We illustrate the application of this algorithm on numerically obtained solutions of the Kuramoto-Sivashinksy partial differential equation.

1 Introduction

In the last few years there has been significant interest in finding low-dimensional representations of data that lies on a manifold [2, 6, 9, 13, 15, 17]. Each of the methods proposed above have certain advantages and disadvantages. One desirable feature for an embedding algorithm would be to have an isometric embedding of the data manifold. None of the prior techniques explicitly attempts to numerically satisfy this optimization

*LA-UR-06-3495

†email: dreisigm@lanl.gov
‡email: kirby@math.colostate.edu

criterion for a general Riemannian manifold. Here we define such an optimization criterion and propose an algorithm that provides a pseudo-isometric embedding of data sampled from a Riemannian manifold.

The problem of finding low-dimensional embeddings of manifolds has a long history. The pursuit of solutions to this problem has led to some of the most beautiful and surprising results in all of mathematics. In 1944 Whitney proved that an m-dimensional manifold has an embedding in $\mathbb{R}^{2m+1}$ [18]. The method of proof was by finding an acceptable projection direction that reduces the dimension of the manifold embedding by one. Whitney's theorem provides a blueprint for a numerical smooth embedding algorithm that was proposed in [2] to find low-dimensional representations of data on manifolds. We will provide a simple modification to this technique that will make the embedding closer to isometric.

2 Whitney's and Nash's Embedding Algorithms

By an embedding of an m-dimensional manifold $\mathcal{U}$ into $\mathbb{R}^n$, $n > m$, we mean that $\mathcal{U} \subset \mathbb{R}^n$ locally looks like $\mathbb{R}^m$ at every point in $\mathcal{U}$. From a practical viewpoint, we would like the embedding dimension n to be as small as possible. This corresponds to compressing the data that lie on $\mathcal{U}$. In addition to compressing the data as much as possible, we would also like to distort the intrinsic distances between points on $\mathcal{U}$ as little as possible. If we retain all of the distances perfectly in our low-dimensional embedding, then we have an isometric embedding.

If we don't care about distorting our data manifold as we compress the ambient embedding dimension, Whitney's theorem gives us an upper bound on the maximum possible embedding dimension [8].

Theorem 2.1 (Whitney's Easy Embedding Theorem) *Let $\mathcal{U}$ be a compact Hausdorff $\mathcal{C}^r$ m-dimensional manifold, $2 \leq r \leq \infty$. Then there is a $\mathcal{C}^r$ embedding of $\mathcal{U}$ in $\mathbb{R}^{2m+1}$.*

The method of proof is to find a projection direction that does not lie parallel to any vector in $\bar{\Sigma}$, where

$$\Sigma \doteq \left\{ \frac{x-y}{\|x-y\|_2} : (x,y) \in \mathcal{U} \times \mathcal{U}, x \neq y \right\}. \tag{1}$$

The set Σ is all of the secant vectors for our manifold $\mathcal{U}$, and $\bar{\Sigma}$ contains, in addition, all of the tangent vectors to $\mathcal{U}$. In [2], it was shown that Whitney's theorem gives a reasonable algorithm for finding low-dimensional embeddings of data manifolds. Further techniques for finding a 'good' projection direction were developed in [3].

Now, suppose we desire an isometric embedding, so our compressed data is 'faithful' to the original data. Here, we can use Nash's $\mathcal{C}^1$ isometric embedding theorem [12].

Theorem 2.2 (Nash's $\mathcal{C}^1$ Isometric Embedding Theorem) *Let $\mathcal{U}$ be a Riemannian m-dimensional manifold. Then there is a $\mathcal{C}^1$ isometric embedding of $\mathcal{U}$ in $\mathbb{R}^{2m+1}$.*

Notice that Nash's theorem gives an upper bound on the embedding dimension that is the same as that in Whitney's theorem. One would correctly suspect that finding the isometric embedding would be a more difficult task than just finding an embedding. In fact, the method of proof in Nash's theorem does not lend itself to numerical implementation. There is a 'spiraling' procedure in Nash's algorithm that requires wrapping very high frequency, low amplitude helixes around an initial embedding. In this way one can 'stretch out' a non-isometric embedding.

So, we know that it is fairly easy to find a low-dimensional embedding of a data manifold. However, improving on this embedding may be a difficult proposition. One idea would be to 'stretch out' the initial embedding provided by an application of Whitney's theorem. Of course we can't do this using Nash's method but, we can do a simple stretching of the embedding. This is the main idea of our algorithm.

3 A Pseudo-Isometric Embedding Algorithm

Here we present a new method for finding the best projection direction to reduce the embedding dimension by one. By 'best direction' we mean finding a direction $\mathbf{p}$ such that the maximum of the inner products $\mathbf{s}_i \cdot \mathbf{p}$ is minimized for all $\mathbf{s}_i \in \Sigma$, the set of unit secants. (In practice we only have access to a subset of Σ because our manifold is sampled.) Let $S \doteq [\mathbf{s}_1|\cdots|\mathbf{s}_j]$ be all of our available unit secant vectors. Then the problem of finding the best projection direction is stated as finding

$$\mathbf{p}^* \quad = \quad \min_{\|\mathbf{p}\|_2=1} \|S^T\mathbf{p}\|_\infty. \tag{2}$$

Now let $\mathbf{q}^T = [\mathbf{p}^T \epsilon]$. Then (2) is equivalent to solving

$$\begin{aligned} \text{minimize} \quad &: \quad \epsilon \\ \text{subject to} \quad &: \quad \begin{bmatrix} S^T & -\mathbf{1} \end{bmatrix} \mathbf{q} \leq 0 \\ & \quad\;\; \mathbf{p}^T\mathbf{p} - 1 = 0. \end{aligned} \tag{3}$$

If we relax the normalization condition on $\mathbf{p}$, then our problem becomes

$$\begin{aligned} \text{minimize} \quad &: \quad -\mathbf{p}^T\mathbf{p} \\ \text{subject to} \quad &: \quad S^T\mathbf{p} \leq 1. \end{aligned} \tag{4}$$

After finding a solution $\hat{\mathbf{p}}$ to (4), we can let

$$\mathbf{p}^* \quad = \quad \frac{\hat{\mathbf{p}}}{\|\mathbf{p}\|_2} \text{ and} \tag{5}$$

$$\epsilon^* \quad = \quad \frac{1}{\|\mathbf{p}\|_2}, \tag{6}$$

which will be solutions to (3). A method for finding a global solution to (4) is presented in [1]. A direct search method for solving (2) directly is given in [7].

Having found our best projection direction $\mathbf{p}^*$, we can now reduce the embedding dimension by one. The new embedding will generally be 'short', i.e., the metric at each point will be smaller than the original metric. However, we can dilate the new embedding space to try to make the new embedding as close to isometric as possible. Now we need to find an optimal dilation. To do this we will pay close attention to how the projected tangent planes are distorted. The way we measure the distance between two points on an embedded manifold $\mathcal{U}$ is by taking infinitesimal steps. These steps are taken in the tangent planes of the points on a manifold. During a compression of the data, e.g., a Whitney projection, we can distort the tangent planes of our manifold. This will in turn distort the distances measured on the manifold. The idea is to undo, as much as is computationally possible, this tangent plane distortion which in turn makes our compressed manifold as close to isometric as possible.

Let the (approximate) tangent vectors to our initial collection of data in the ambient space $E = [e^{(1)}, \ldots, e^{(n)}]$ be given by

$$T \quad = \quad \left[t^{(1)} | \cdots | t^{(n)} \right], \tag{7}$$

where $t^{(i)} = [t_1^{(i)}, \ldots, t_m^{(i)}]$. The desired metric at each data point is $g^{(i)} = t^{(i)T} t^{(i)}$, and the new tangent vectors are given by $\hat{t}^{(i)} = (I - \mathbf{p}^* \mathbf{p}^{*T}) t^{(i)}$. Then finding the optimal dilation can be stated as the semidefinite program (SDP) [16]

$$\min_{M \geq 0} \max_{i,j,k} |g_{j,k}^{(i)} - \hat{t}_j^{(i)T} M \hat{t}_k^{(i)}|, \tag{8}$$

with $j \leq k = 1, \ldots, m$. Letting $T_{j,k}^{(i)} = \hat{t}_j^{(i)} \hat{t}_k^{(i)T}$, (8) becomes

$$\begin{array}{rcl} \text{minimize} & : & \iota \\ \text{subject to} & : & M \bullet T_{j,k}^{(i)} - \iota + z_{j,k}^{(i)} = g_{j,k}^{(i)} \\ & & -M \bullet T_{j,k}^{(i)} - \iota + \zeta_{j,k}^{(i)} = -g_{j,k}^{(i)} \\ & & z_{j,k}^{(i)},\ \zeta_{j,k}^{(i)},\ \iota \geq 0 \\ & & M \geq 0, \end{array} \tag{9}$$

where $\bullet$ denotes the matrix inner product

$$A \bullet B \quad = \quad \sum_{r,c} a_{r,c} b_{r,c}. \tag{10}$$

Here ι is an 'isometric index' and is positively correlated with the departure from isometry, and satisfies $0 \leq \iota < 1$. Hence, a lower value of ι means we have a closer-to-isometric embedding. Solving (9) for M, we can dilate the reduced dimensional 'short' embedding $\hat{E}$ to find a pseudo-isometric embedding $\tilde{E}$ by letting

$$\tilde{E} \quad = \quad M^{1/2} \hat{E}. \tag{11}$$

Note that $\tilde{E}$ generally will not be a 'short' embedding. What we are doing in (11) is stretching the ambient space in order to stretch out the tangent spaces of $\hat{E}$. In this way we are reducing the distortion of the tangent spaces introduced by the Whitney projection algorithm in (2).

The SDP in (9) will typically be too large to practically solve. However, we can take a sample of the data points and their corresponding tangent vectors and solve the reduced problem. This 'landmarks' version seems reasonable given we only need a sufficiently dense sampling of the tangent vectors to find M. Also, the process can be made 'blind' by choosing the smallest secants at each data point to be approximations to tangent vectors. Alternately, we can use the SVD based method in [4, 5] to find orthonormal bases for the tangent spaces, in which case $g^{(i)} = I$ in (9) since $t^{(i)T} t^{(i)} = I$. The latter technique is the one used in this paper for the numerical experiments.

4 Consistency Test

When using the SVD procedure for finding the tangent space, we need to show that the spans of the projection of the high-dimensional tangent spaces correspond to the spans of the tangent spaces found from using the SVD procedure on the reduced dimensional embedding. For the data point $\mathbf{e}^{(i)}$ in the high-dimensional embedding, let $N^{(i)}$ be the matrix of nearest neighbors used to find the tangent space. So $\mathbf{e}^{(i)}$ is a column of E, and $N^{(i)}$ consists of the columns of E that are closest to $\mathbf{e}^{(i)}$. Then, following the procedure in [4, 5], we center the matrix $N^{(i)}$ to find a new matrix $D^{(i)}$ given by

$$D^{(i)} = N^{(i)} - \mathbf{e}^{(i)} \mathbf{1}^T, \tag{12}$$

where $\mathbf{1}^T = [1 \cdots 1]$. Let the SVD of $D^{(i)}$ be given by

$$D^{(i)} = \left[U_1^{(i)} \, U_2^{(i)} \right] \begin{bmatrix} S_1^{(i)} & 0 \\ 0 & S_2^{(i)} \end{bmatrix} \begin{bmatrix} V_1^{(i)T} \\ V_2^{(i)T} \end{bmatrix}. \tag{13}$$

All of the significant singular values will be in $S_1^{(i)}$. If the manifold is m-dimensional, then $S_1^{(i)}$ will be m-by-m. $S_2^{(i)}$ will contain only insignificant eigenvalues under some 'noise' level. These will generally be a few orders of magnitude less than the singular values in $S_1^{(i)}$. Here we assume that the elements in $N^{(i)}$ are close enough to $\mathbf{e}^{(i)}$ such that the effects of the curvature of the manifold can be ignored. It follows then that the columns of $U_1^{(i)}$ give us an orthonormal basis for the tangent space of $\mathbf{e}^{(i)}$.

Let P be the projection operator found from the Whitney algorithm. Then we need to consider

$$\begin{aligned} P^T D^{(i)} &= P^T U_1^{(i)} S_1^{(i)} V_1^{(i)T} + P^T U_2^{(i)} S_2^{(i)} V_2^{(i)T} \\ &= \tilde{D}_1^{(i)} + \tilde{D}_2^{(i)}. \end{aligned} \tag{14}$$

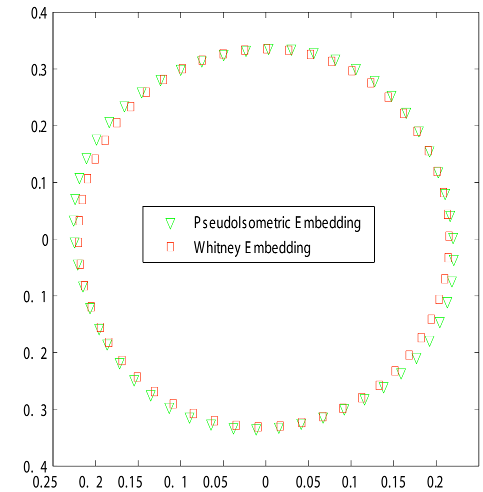

Figure 1: The embedding of the circle in $\mathbb{R}^2$ of the $\alpha = 84$ K-S data set.

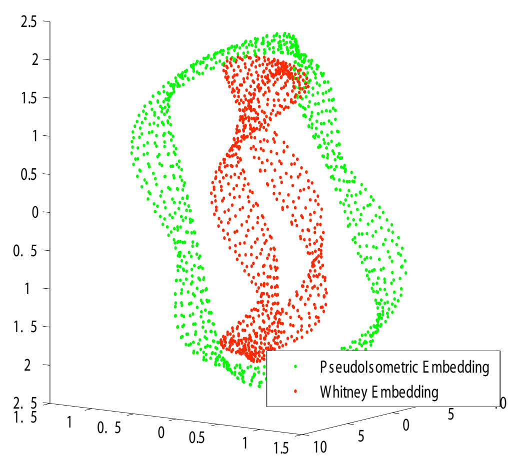

Figure 2: The embedding of the torus in $\mathbb{R}^3$ associated with $\alpha = 87$.

The span of the projected high-dimensional tangent space is given by the right singular vectors of $\tilde{D}_1^{(i)}$. So generally the span of the tangent space calculated from the reduced dimensional embedding will be perturbed from the span of the projected high-dimensional tangent space [14]. The effects of perturbation will depend on how much the spectrum of $\tilde{D}_1^{(i)}$ is separated from the spectrum of $\tilde{D}_2^{(i)}$. As long as we do not significantly collapse any tangent vector, this separation should be significant, and the resulting perturbation minor. This should typically be the case. The reasoning is that a significant collapse of a tangent vector would indicate a projection into a dimension that is too small for an embedding of the manifold. However, it is possible that the projection operator from the Whitney algorithm will significantly increase the local curvature, resulting in a breakdown of the SVD procedure in the reduced dimensional space. This can be mitigated by increasing the density of our sampling of the manifold. In the continuum limit, the problem disappears.

5 Examples

Here we will examine the various algorithms on two data sets. These are taken from the Kuramoto-Sivashinsky (K-S) equation

$$u_t + 4u_{xxxx} + \alpha \left(u_{xx} + \frac{1}{2}(u_x)^2 \right) \quad = \quad 0, \tag{15}$$

with $\alpha = 84$ (a limit-cycle) and $\alpha = 87$ (a torus). The data consists of numerical solutions of the PDE generated by a Fourier-Galerkin method as described in [10, 11]. The simulations used here employed 10 complex Fourier modes and hence all of the data is initially in the ambient space $\mathbb{R}^{20}$. The pseudo-isometric Whitney method was run 'blind' using the SVD procedure in [4, 5] to find the tangent spaces. Thus, we are trying to find an optimal dilation matrix M in (9) to make all of the projected tangent spaces' inner products as close to the identity matrix as possible.

For the $\alpha = 84$ case, we have 54 samples from the limit cycle. The initial "good" (but not isometric) projection was into $\mathbb{R}^2$ and this was improved by determining the matrix M as defined in Equation (8). The actual embedding of this data is show in Figure 1. Note, that, as expected, the embeddings are geometrically closed curves. The errors associated with the metric for the Whitney and pseudo-isometric Whitney algorithms are shown in Figure 3. The eigenvalues of M in this case were 1.0558 and 1.0163, and $\iota = 0.0246$ in (9). For the $\alpha = 87$ case, we have 951 samples from the torus. The samples were on a one dimensional 'thread' that 'looped' around the torus. Here the initial embedding was into $\mathbb{R}^3$. The resulting pseudo-isometric and Whitney embeddings are shown in Figure 2. Although it is not easy to see, they each have the appearance of a flattened torus. A summary of the errors in the metric for this experiment are presented in Figure 4. The eigenvalues of M in this case were 0.2585, 1.3268 and 6.8937, and $\iota = 0.9860$ in (9).

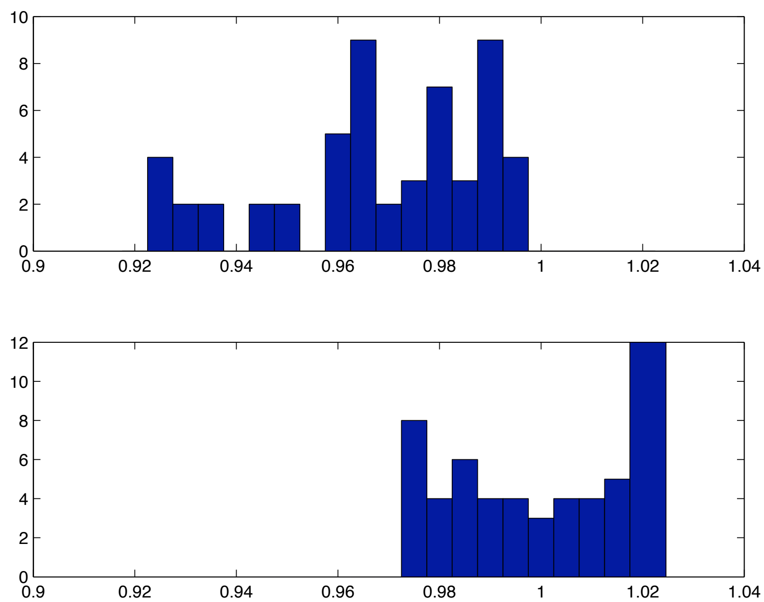

Figure 3: The lengths of the tangent vectors for embeddings in $\mathbb{R}^2$ of the $\alpha = 84$ K-S data set. The upper histogram is for the Whitney algorithm. The bottom histogram is for the pseudo-isometric Whitney algorithm. We want the lengths to be as close to 1 as possible. Note that the pseudo-isometric Whitney algorithm embedding is no longer short.

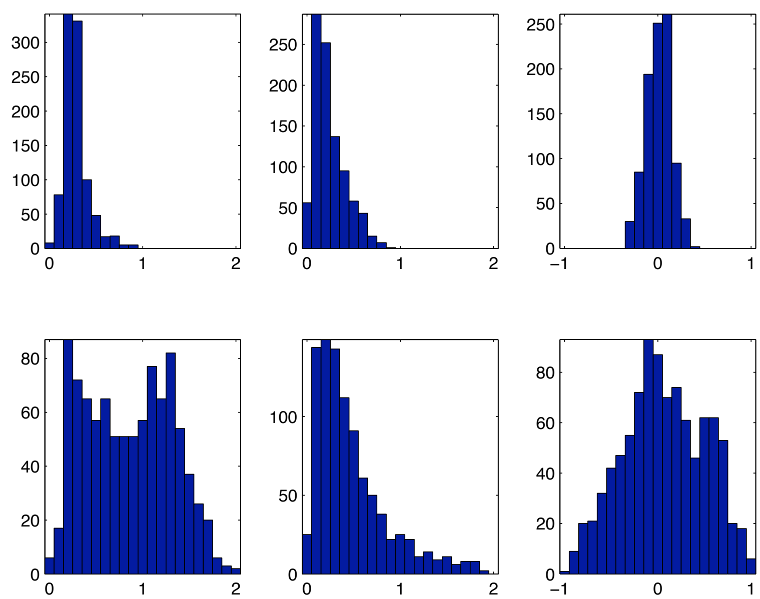

Figure 4: The inner products of the tangent vectors for embeddings in $\mathbb{R}^3$ of the $\alpha = 87$ K-S data set. The upper plots are for the Whitney algorithm. The bottom plots are for the pseudo-isometric Whitney algorithm. From left to right are $t_1^{(i)} \cdot t_1^{(i)}$, $t_2^{(i)} \cdot t_2^{(i)}$, and $t_1^{(i)} \cdot t_2^{(i)}$. We want these to be as close to 1, 1 and 0 as possible, respectively.

6 Discussion

We have introduced a new quadratic program to find the optimal projection direction to be used in Whitney's algorithm. Additionally, we have shown how to modify this embedding to be closer to isometric via a semidefinite program. Both of these algorithms provide reasonable embeddings of high-dimensional data. The main extension of the pseudo-isometric Whitney algorithm would be to develop a 'landmarks' version in order to increase the SDP's performance.

One of the main advantages of these algorithms is the clear analytic limit. Indeed, they are guaranteed to give embeddings into at most $\mathbb{R}^{2m+1}$ for any m-dimensional manifold. The previous algorithms in [2, 6, 9, 13, 15, 17] lack this generality. Additionally, the lowest embedding dimension is easily recognized for a reasonably densely sampled manifold since at least one of the secants available will be significantly collapsed when we try to project into a dimension that is too low.

Other isometric embedding algorithms have been proposed. The Hessian Local Linear Embedding (HLLE) algorithm in [6] and the Isomap algorithm in [15] are restricted to the cases where the entire manifold is isometric to a region of $\mathbb{R}^m$. Theoretically, this means that data that lies on a circle, a sphere or a torus, for example, cannot be isometrically embedded by these algorithms. The Semi-Definite Embedding (SDE) algorithm in [17] does not actually provide embeddings since it can change the topology of the data. For example, it can untie knots. Also, the standard and pseudo-isometric Whitney embedding algorithms give us a measure of the maximum departure from isometry: the ϵ in (3) and the ι in (9).

7 Acknowledgments

Funding was provided in part by the IC Postdoctoral Research Fellowship Program and by the DOD-USAF-Office of Scientific Research under contract FA9550-04-1-0094. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IC Postdoctoral Program or the DOD-USAF-Office of Scientific Research.

References

[1] I. Bomze and G. Danninger. A global optimization algorithm for concave quadratic programming problems. *SIAM Journal on Optimization*, 3:826–842, 1993.

[2] D. Broomhead and M. Kirby. A new approach to dimensionality reduction: Theory and applications. *SIAM Journal of Applied Mathematics*, 60:2114–2142, 2000.

[3] D. Broomhead and M. Kirby. Dimensionality reduction using secant-based projection methods: The induced dynamics in projected systems. *Nonlinear Dynamics*, 41:47–67, 2005.

[4] D. S. Broomhead, R. Indik, A. C. Newell, and D. A. Rand. Local adaptive Galerkin bases for large-dimensional dynamical systems. *Nonlinearity*, 4:159–197, 1991.

[5] D. S. Broomhead, R. Jones, and G. P. King. Topological dimension and local coordinates from time series data. *Journal of Physics A*, 20:L563–L569, 1987.

[6] D. L. Donoho and C. Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences (U.S.A.)*, 100:5591–5596, 2003.

[7] D. W. Dreisigmeyer. Direct search algorithms over Riemannian manifolds. Submitted to SIOPT.

[8] M. W. Hirsch. *Differential Topology*. Springer, 1976.

[9] I. Hotz. Isometric embedding by surface reconstruction from distances. In *Proceedings of the Conference on Visualization 2002*, pages 251 – 258, 2002.

[10] M. Kirby and D. Armbruster. Reconstructing phase-space from PDE simulations. *Z. angew. Math. Phys.*, 43:999–1022, 1992.

[11] M. Kirby and R. Miranda. Nonlinear reduction of high-dimensional dynamical systems via neural networks. *Phys. Rev. Letters*, 72(12):1822–1825, 1994.

[12] J. Nash. $\mathcal{C}^1$ isometric embeddings. *Annals of Mathematics*, 60:383–396, 1954.

[13] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by local linear embedding. *Science*, 290:2323–2326, 2000.

[14] G. Stewart. Perturbation theory for the singular value decomposition. In R. Vaccaro, editor, *SVD and Signal Processing, II: Algorithms, Analysis and applications*, pages 99–109. Elsevier, 1991.

[15] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.

[16] L. Vandenberghe and S. Boyd. Semidefinite programming. *Siam Review*, 38:49–95, 1996.

[17] K. Q. Weinberger, F. Sha, and L. K. Saul. Learning a kernel matrix for nonlinear dimensionality reduction. In *Proceedings of the* 21^{st} *International Conference on Machine Learning*, page 106, 2004.

[18] H. Whitney. The self-intersections of a smooth n-manifold in $2n$-space. *Annals of Mathematics*, 45:220–246, 1944.